LESIONS DETECTION ON 3D BRAIN MRI USING TRIMMMED LIKELIHOOD ESTIMATOR AND PROBABILISTIC ATLAS

S. Bricq[*]*, Ch. Collet*

Strasbourg University
LSIIT - UMR CNRS 7005
Pole API, Bd S. Brant,
F-67412 Illkirch - France

J.-P. Armspach

Strasbourg University
LINC - UMR CNRS 7191
4 rue Kirschleger
F-67205 Strasbourg - France

ABSTRACT

In this paper, we present a new automatic robust algorithm to segment multimodal brain MR images with Multiple Sclerosis (MS) lesions. The method performs tissue classification using a Hidden Markov Chain (HMC) model and detects MS lesions as outliers to the model. For this aim, we use the Trimmed Likelihood Estimator (TLE) to extract outliers. Furthermore, neighborhood information is included using the HMC model and we propose to incorporate *a priori* information brought by a probabilistic atlas. Tests on Brainweb images with MS lesions have been carried out to validate this approach.

Index Terms— Image segmentation, Hidden Markov models, robustness, Magnetic Resonance Imaging

1. INTRODUCTION

Multiple Sclerosis (MS) is a disorder of the central nervous system. To better understand this disease and to quantify its evolution, magnetic resonance imaging (MRI) is increasingly used nowadays. Nevertheless, manual delineation of lesion by human experts is a time-consuming process and is prone to intra- and inter-observer variability, which deteriorates the significance of the resulting segmentation analysis. Therefore fully automated and reproducible methods are required to segment MS lesions in multimodal MR sequences. MS lesions are often detected as voxels that are not well explained by a statistical model for normal brain MR images.

In [1], Schroeter *et al.* use Gaussian mixtures to model the presence of different components within each voxel, and to robustify the estimation scheme, they add a class of outliers with a uniform distribution corresponding to lesions. In [2], Van Leemput *et al.* introduce weights reflecting the degree of typicality of each voxel. Their method also includes neighborhood information using a Potts model. In this paper, we propose to keep neighborhood information during the inference process by using a Hidden Markov Chain model taking into account *a priori* information brought by a probabilistic atlas, and detecting outliers using the Trimmed Likelihood Estimator.

This paper is organized as follows: next section introduces the Hidden Markov Chain model and explains how information brought by a probabilistic atlas is incorporated to help the segmentation process. Section 3 presents the Trimmed Likelihood Estimator (TLE) used to detect outliers corresponding to ME lesions. In section 4, we apply such estimator to the HMC model for MS lesion detection: results obtained on Brainweb images are shown in section 5. Finally in section 6, conclusions are drawn and future developments are suggested.

2. ROBUST HIDDEN MARKOV CHAIN SEGMENTATION USING PROBABILISTIC ATLAS

To segment Brain MRI, we propose to use Hidden Markov Chains (HMC) by using a 3D Hilbert-Peano scan of the data cube [3]. HMC is a method based on neighborhood information which has been widely used to segment 2D images (see e.g. [4]). Neighborhood information is included in the HMC model. The interest of Markov Chain methods for image segmentation compared to 3D Markov Random Field (MRF) models is that being based on 1D modeling, they result in lower computing costs with similar results. Contrary to MRF, the neighboring information is partially translated in the chain: two neighbors in the chain are neighbors in the grid, but two neighbors in the cube can be far away in the chain. However, due to strong correlation within the data cube, this scan will weakly influence the segmentation results. The first step of segmentation algorithms based on HMC consists in transforming the image into a vector[3]. Once all the processing has been carried out on the vector, the inverse transformation is applied on the segmented chain to obtain the final segmented image.

Let us now consider two sequences of random variables $X = (X_n)_{n \in S}$ the hidden process, and $Y = (Y_n)_{n \in S}$ the observed one, with S the finite set corresponding to the N

[*]Thanks to Region of Alsace for funding.

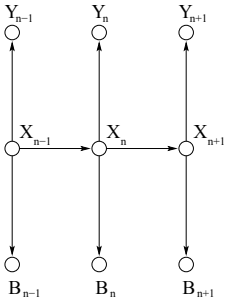


Fig. 1. *Dependency graph of Hidden Markov Chain with atlas information*

voxels of the image. Each X_n takes its value in a finite set of K classes $\Omega = \{\omega_1, \ldots, \omega_K\}$ and each Y_n takes its value in $\mathbb{R}$. For this application, we have $K = 3$ classes $\Omega = \{WM, GM, CSF\}$ where WM, GM and CSF denote respectively white matter, gray matter and cerebrospinal fluid. X is a Markov Chain if $P(X_{n+1} = \omega_{k_{n+1}} | X_n = \omega_{k_n}, \ldots, X_1 = \omega_{k_1}) = P(X_{n+1} = \omega_{k_{n+1}} | X_n = \omega_{k_n})$. Thus X will be determined by the initial distribution $\pi_k = P(X_1 = \omega_k)$ and the transition matrix $a_{kl}^n = P(X_{n+1} = \omega_l | X_n = \omega_k)$. We assume the homogeneity of the Markov Chain which means that the transition matrix is independent of the location n: $a_{kl}^n = a_{kl}$, for $1 \leq n < N$.

The likelihood $f_k(y_n; \theta) = P(Y_n = y_n | X_n = \omega_k)$ of the observation y_n conditionally to $X_n = \omega_k$ is assumed to be a Gaussian density with mean $\mu_k \in \{\mu_1, \ldots, \mu_K\}$ and variance $\sigma_k^2 = \{\sigma_1^2, \ldots, \sigma_K^2\}$. These parameters are clustered in θ. *A priori* information brought by a probabilistic atlas is introduced in the model to drive the segmentation process. This atlas derived from 31 normal brains which were registrated using a non-rigid transformation [5] and segmented using a HMC model [6]. Then these different segmentations were averaged to obtain the atlas. This atlas contains probability information about the expected location of WM, GM, and CSF. The different probabilities in each voxel n calculated during the HMC algorithm were multiply by the prior probability $b_n(k)$ of this voxel to belong to class k given by the atlas in the HMC modeling. The dependency graph of a HMC is presented in Fig. 1. One of the interests of Hidden Markov Chain is the possibility of computing exactly the posterior marginals at each location and to obtain a labeling $\hat{x}$ of the image by using the posterior probability [7]:

$$\hat{x}_n = \arg\max_{\omega_k \in \Omega} P(X_n = \omega_k | Y = y) \quad (1)$$

$$= \arg\max_{\omega_k \in \Omega} \alpha_n(k) \beta_n(k) \quad (2)$$

with $\alpha_n(k) = P(X_n = \omega_k, Y_1, \ldots, Y_n)$ forward probability and $\beta_n(k) = P(Y_{n+1}, \ldots, Y_N | X_n)$ backward probability. These probabilities can be computed recursively [8]. This recursive computation is detailed in Sec 4 in the robust case.

3. TRIMMED LIKELIHOOD ESTIMATOR

We detect MS lesions as outliers toward statistical model of normal brain images. To extract these outliers and to estimate the parameters of the different classes in a robust way, the Trimmed Likelihood Estimator (TLE) was used. The TLE was introduced in [9] and developed to estimate mixture of multivariate normals and generalized linear models in a robust way [10]. The main idea lies in removing the $n - h$ observations whose values would be highly unlikely to occur if the fitted model was true. The optimization scheme used to compute this estimator derives from the optimization scheme of the Least Trimmed Squares (LTS) estimators of Rousseeuw and Leroy [11]. This algorithm was used to segment brain MRI by Aït-Ali in the frame of Gaussian mixtures without including neighborhood and atlas information [12].

3.1. Trimmed Likelihood Estimator

The Trimmed Likelihood Estimator (TLE) [13] is defined as:

$$\hat{\theta}_{TLE} = \arg\min_{\theta \in \Theta^p} \sum_{i=1}^{h} \psi(y_{\nu(i)}; \theta) \quad (3)$$

where for a fixed θ, $\psi(y_{\nu(1)}; \theta) \leq \psi(y_{\nu(2)}; \theta) \leq \ldots \leq \psi(y_{\nu(n)}; \theta)$, $\psi(y_i; \theta) = -\log f(y_i; \theta)$, $y_i \in \mathbb{R}^q$ for $i = 1, \ldots, n$ are i.i.d observations with probability density $f(y, \theta)$ depending on an unknown parameter $\theta \in \Theta^p \subset \mathbb{R}^p$. $\nu = (\nu(1), \ldots, \nu(n))$ is the corresponding permutation of the indices, which depends on θ, and h is the trimming parameter.

$$\hat{\theta}_{TLE} = \arg\max_{\theta \in \Theta^p} \prod_{i=1}^{h} f(y_{\nu(i)}; \theta) \quad (4)$$

General conditions for the existence of a solution of (Eq. 3) are proved in [14]. Convergence and asymptotic properties are studied in [15].

3.2. FAST-TLE algorithm

In [16], Neykov and Müller develop a fast iterative algorithm for derivation of the TLE. This FAST-TLE algorithm can be described as follows: given $H^{old} = \{y_{j_1}, \ldots, y_{j_n}\} \subset \{y_1, \ldots, y_n\}$,

- Compute $\hat{\theta}^{old} := MLE$ (Maximum Likelihood Estimator) based on H^{old}.
- Define $Q^{old} = \sum_{i=1}^{k} \psi(y_{j_i}, \hat{\theta}^{old})$.
- Sort $\psi(y_i, \hat{\theta}^{old})$ for $i = 1, \ldots, n$ in ascending order: $\psi(y_{\nu(i)}, \hat{\theta}^{old}) \leq \psi(y_{\nu(i+1)}, \hat{\theta}^{old})$ and get the permutation $\nu = (\nu(1), \ldots, \nu(n))$.
- Define $H^{new} = \{y_{\nu(1)}, \ldots, y_{\nu(n)}\}$.

- Compute $\hat{\theta}^{new} := MLE$ based on H^{new}.
- Define $Q^{new} = \sum_{i=1}^{k} \psi(y_{\nu(i)}, \hat{\theta}^{new})$.

4. PROPOSED FRAMEWORK

To estimate parameters in a robust way and to detect lesions, we thus adapt the FAST-TLE algorithm presented in Sec. 3.2 to the HMC model. We use the following notations: $f_k(y_n;\theta) = P(Y_n = y_n|X_n = \omega_k)$ denotes the likelihood of the observation y_n conditionally to $X_n = \omega_k$ and $b_n(k)$ represents the prior probability of voxel n to belong to class k given by the atlas B. This leads to:

1. Compute $\hat{\theta}^{(p-1)} := MLE$ using the Expectation-Maximization (EM) algorithm [17], based on the whole dataset;
2. Sort residus $r_n = -\log f(y_n, b_n; \hat{\theta}^{(p-1)}) = -\log P(Y_n = y_n, B_n, \hat{\theta}^{(p-1)})$ for $n = 1, \ldots, N$ with:

$$\begin{aligned} P(Y_n = y_n, B_n, \hat{\theta}^{(p-1)}) &= \sum_{\omega_k} P(Y_n = y_n, B_n, X_n = \omega_k, \hat{\theta}^{(p-1)}) \\ &= \sum_{\omega_k} P(X_n = \omega_k) b_n(k) f_k(y_n; \hat{\theta}^{(p-1)}) \end{aligned} \quad (5)$$

3. Define $H^{(p)} = \{y_{\nu(1)}, ..., y_{\nu(h)}\}$ the subset containing the h vectors with the lowest residus for $\hat{\theta}^{(p-1)}$;
4. Compute $\hat{\theta}^{(p)} := MLE$ using EM, based on $H^{(p)}$. We assign the likelihood of data considered as outliers to one, i.e. $f_k(y_n) = 1, \forall k$ in the HMC process. On the location where the data is considered as an outlier, only prior distribution takes place in the labeling process. Calculation of the different probabilities becomes:
   - Forward probabilities:
     - $\alpha_1(k) = \pi_k f_k(y_1; \hat{\theta}^{(p)}) b_1(k)$
     - $\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) a_{lk} f_k(y_n; \hat{\theta}^{(p)}) b_n(k)$ with $f_k(y_n, \hat{\theta}^{(p)}) = 1$ if y_n is considered as an outlier.
   - Backward probabilities:
     - $\beta_N(k) = 1$
     - $\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) a_{kl} f_l(y_{n+1}; \hat{\theta}^{(p)}) b_{n+1}(l)$ with $f_l(y_{n+1}, \hat{\theta}^{(p)}) = 1$ if y_{n+1} is considered as an outlier.
   - *a posteriori* joint probabilities:

$$\begin{aligned} \xi_n(i,j) &= P(X_n = \omega_i, X_{n+1} = \omega_j | Y = y, B) \\ &= \frac{\alpha_{n-1}(j) a_{ji} f_i(y_n, \hat{\theta}^{(p)}) b_n(i) \beta_n(i)}{\sum_k \alpha_n(k)} \end{aligned}$$

   - *a posteriori* marginal probabilities: $\gamma_n(i) = P(X_1 = \omega_i | Y_1, ..., Y_N) = \frac{\alpha_n(i)\beta_n(i)}{\sum_j \alpha_N(j)}$
   - $\mu_i = \frac{\sum_{n_1} \gamma_{n_1}(i) y_{n_1}}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$.
   - $\sigma_i = \frac{\sum_{n_1} \gamma_{n_1}(i)(y_{n_1} - \mu_i)(y_{n_1} - \mu_i)^t}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$.
5. Back to step 2 until convergence of $H^{(p)}$.

The main drawback of this approach is that the trimming parameter h representing the percentage of voxels used to estimate the parameters has to be fixed by the user. To carry out this problem, we propose to use an adaptative trimming parameter and a threshold s for the probability $P(Y_n = y_n, B_n, \theta)$ (Eq. 5). At each iteration, the voxels for which the probability $P(Y_n = y_n, B_n, \theta)$ is lower than the threshold s are considered as outliers to the model and not included in HMC parameter estimation. In this case, the trimming parameter h will change at each iteration.

Outlier voxels also occur outside MS lesions, especially in the CSF class. Thus to remove these outliers which are not MS lesions, a post-processing step was added to our algorithm. Outliers for which the prior probability of CSF given by the atlas is higher than 0.5 were removed and lesions with a small volume ($3mm^3$) were excluded.

5. VALIDATION

We applied the robust HMC model presented in previous section to brain MRI segmentation. This method has been tested on the Brainweb database[1] [18] which offers a large amount of different phantoms of multimodal MR brain images with MS lesions with different noise and non-uniformity levels. From these phantoms, the tissue classification in WM, GM, CSF and MS lesion is known. To evaluate the performance of our algorithm, we use the Kappa index (KI):

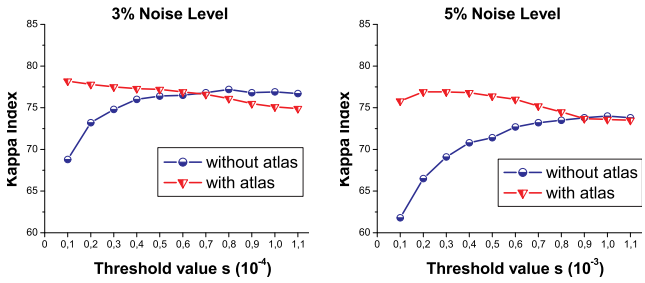
$$KI = 2\frac{SEG \bigcap GT}{SEG + GT} \quad (6)$$

where GT stands for the ground truth and SEG for the segmentation obtained.

The method was tested on T1/T2 images with 3 and 5% of noise and 20% inhomogeneity level for different values of the threshold s. Comparisons of the results for lesions segmentation with and without atlas information are presented in Fig. 2 for 3 and 5% of noise. The highest Kappa index obtained for 3% (respectively 5%) of noise without atlas information is 77.2% (respectively 74%), whereas the highest Kappa index obtained for 3% (respectively 5%) of noise with atlas information has a value of 78.2% (respectively 76.9%). Segmentation obtained using atlas information yields the best results.

[1] http://www.bic.mni.mcgill.ca/brainweb/

Moreover the threshold used to obtain the highest Kappa index is lower using atlas information. This means that lesions are better and easier detected using atlas information.



(a) (b)

Fig. 2. Kappa index for lesions obtained on T1/T2 Brainweb images for different values of the threshold parameter s with and without atlas information. (a) and (b) correspond respectively to the results obtained for 3% and 5% of noise. Lesions are better detected using atlas information.

6. CONCLUSION AND OUTLOOK

We have described a robust framework for tissue classification of multimodal brain MR images and MS lesions detection. Hidden Markov Chains were used to include neighborhood information in the model. This spatial regularization is required to overcome the disturbance added during the MRI formation. Moreover *a priori* information was introduced using a probabilistic atlas and lesion extraction was carried out using the Trimmed Likelihood Estimator and an adaptative threshold. This model has been validated for lesion detection on 3D multimodal brain phantoms with MS lesions, and will be applied to segment real brain images with MS lesions and compared to manual expert segmentation.

7. REFERENCES

[1] P. Schroeter, J.-M. Vesin, T. Langenberger, and R. Meuli, "Robust parameter estimation of intensity distributions for brain magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 172–186, April 1998.

[2] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 677–688, August 2001.

[3] Y. Bandoh and A. Kamata, "An address generator for a 3-dimensional pseudo-Hilbert scan in acuboid region," in *Proc. of IEEE Int. Conf on Image Process.*, 1999.

[4] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised Classification of Radar Images Using Hidden Markov Chains and Hidden Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 675–686, March 2003.

[5] V. Noblet, C. Heinrich, F. Heitz, and J.P. Armspach, "3-D Deformable Image Registration : A Topology Preservation Scheme Based on Hierarchical Deformation Models and Interval Analysis Optimization," *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 553–566, 2005.

[6] S. Bricq, Ch. Collet, and J.-P. Armspach, "Triplet Markov Chain for 3D MRI brain segmentation using a probabilistic atlas," in *IEEE 2006 International Symposium on Biomedical Imaging, April 6-9, ISBI'06*, 2006.

[7] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian data analysis*, Chapman and Hall - New York, 2005.

[8] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recognition Letters*, vol. 3, no. 6, pp. 369–373, December 1985.

[9] N.M. Neykov and P.N. Neytchev, "A robust alternative of the MLE," *Compstat'90*, pp. 99–100, 1990.

[10] D.L. Vandev and N.M. Neykov, "Robust Maximum Likelihood in the Gaussian Case," in *New Directions in Data Analysis and Robustness*, S. Morgenthaler et al., Ed., Birkhäuser Verlag Basel, Switzerland, 1993, pp. 259–264.

[11] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 1987.

[12] L.S. Aït-ali, S. Prima, P. Hellier, B. Carsin, G. Edan, and C. Barillot, "STREM: a robust multidimensional parametric method to segment MS lesions in MRI," in *MICCAI'2005*, J. Duncan and G. Gerig, Eds., Palm Springs, USA, October 2005, vol. 3749 of *Lecture Notes in Computer Science*, pp. 409–416, Springer.

[13] A.S. Hadi and A. Luceno, "Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms," *Computational Statistics and Data Analysis*, vol. 25, pp. 251–272, 1997.

[14] R. Dimova and N.M. Neykov, "Generalized d-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications," *Theory and applications of recent robust methods*, 2004.

[15] P. Čížek, "Robust estimation in nonlinear regression and limited dependent variable models," *EconPapers*, 2002.

[16] N.M. Neykov and C.H. Müller, "Breakdown point and computation of trimmed likelihood estimators in generalized linear models," in *Developments in robust statistics*, R. Dutter, P. Filzmoser, U. Gatter, and P.J. Rousseeuw, Eds., Physica-Verlag, Heidelberg, 2003, pp. 277–286.

[17] M.A. Tanner, *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions*, Springer Verlag, 1993.

[18] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "An Extensible MRI Simulator for Post-Processing Evaluation," *Visualization in Biomedical Computing (VBC'96)*, vol. 1131, pp. 135–140, 1996.